

Contribution of Nutrients and *E. coli* to Surface Water Condition in the Ozarks
I. Using Partial Least Squares Predictions when Standard Regression Assumptions Are Violated

Authors: Maliha S. Nash, Ricardo Lopez
U.S. EPA/LEB

Keywords: PLS, surface water *E. coli*, surface water nutrient, landscape, Ozark

The primary objective of the U.S. EPA's Landscape Ecology research program is investigation of associations between surface water constituents and landscape metrics. Statistically valid predictive models are an important means of expressing these associations. Models developed by using standard multiple or multivariate regression analyses frequently exclude important variables due to problems or restrictions with the available data sets. Common data problems in ecological studies include small sample size, missing values, a large number of predictors, correlated variables, and high noise-to-signal relationships. Partial least square (PLS) regression was developed particularly to deal with the above problems. We present here the application of PLS regression to predicting surface water total phosphorous, total ammonia, and *Escherichia coli* (*E. coli*) from landscape metrics. The amount of variability in surface water constituents explained by each model reflects the composition of the contributing landscape metrics. The landscape–water model developed using PLS explains 59%, 93%, and 81% of the variation in surface water total phosphorous, total ammonia, and *E. coli*, respectively.

Notice: Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy.

Point of Contact:

Maliha Nash
Statistician
U.S. EPA/LEB
944 East Harmon Avenue
Las Vegas, NV 89119-6748
702-798-2201
nash.maliha@epa.gov